PaN-data ODI

Deliverable D3.1

Specification of AAI infrastructure

Grant Agreement Number	RI-283556
Project Title	PaN-data Open Data Infrastructure
Title of Deliverable	Specific requirements for the virtual laboratories
Deliverable Number	D3.1
Lead Beneficiary	PSI
Deliverable Dissemination Level	Public
Deliverable Nature	Report
Contractual Delivery Date	01 Apr 2012 (Month 6)
Actual Delivery Date	29 Aug 2012, rev. 10 Jul 2013

The PaN-data ODI project is partly funded by the European Commission under the 7th Framework Programme, Information Society Technologies, Research Infrastructures.

FIM Architecture Document
PaNdata / WP3

Version 3.1
July 2012, rev 2013-07-11 13:59

1 Introduction

1.1 The CRISP / PaNdata AAI activities

1.1.1 Two Workpackages

Within the last years, Federated Identity management (FIM) at large research facilities has gained strong interest and is presently one of the most discussed IT topics at these facilities. This is reflected by the fact that there are two work packages of two totally decoupled FP7 projects (CRISP/WP16 and PaNdata/WP3) are dealing with this topic in such a way, that there is a strong overlap. The processes defining the work description of these projects and defining the contents of the workpackages were completely independent, so that during this planning phase no synchronization was possible. As soon, however, as both projects were approved, the representatives of these projects started a close collaboration in order to avoid any double actions and, instead, to maximize synergy effects.

The CRISP and PaNdata WPs will both deal with solution of the FIM demands. Concerning the job division, it is proposed that instead of a simple vertical subdivision (which subtopic is handled by which WP) there should be a horizontal subdivision, which is in line also with the respective WP descriptions. According to that, PaNdata/WP3 will care more about implementation of the AAI tool, whereas CRISP/WP16 will care more about further developments.

The partners in PaNdata/WP3 and CRISP/ WP16 represent only a part of the nearly two dozen European photon and neutron facilities currently in operation with even more under construction.

Furthermore, many of the more than 30000 Photon and Neutron facility users perform experiments at different European facilities and a modern FIM tool is urgently needed to guarantee access to these facilities and their common services at the European level. In order to extend the FIM solution outside the P/N-community, (i.e. most users belong to the federation their home institution is part of) connection to other Federations is being developed within CRISP/WP16.

In any case, such a system will be able on the long term only if it is a common system and accepted by all facilities. This aspect of the need for a general solution has always to be kept in mind.

1.1.2 Consequences on the document structures

This subdivision reflects itself in the structure of the CRISP and PaNdata deliverables D16.1 [9] and D3.1 [present work]. Sections 1 (introduction), 2 (Federated Identity Management approaches) and 3 Umbrella description are nearly identical in both architecture documents, the Umbrella implementation / deployment-specific issues are treated in section 4. CRISP / WP16 deals with issues further away from the basic Umbrella services (see section 5).

1.2 The FIM landscape

The increased use of IT tools in academic research and everyday life in general is accompanied by an increased need for identity services. Well known are the developments in the banking and government administration sectors. But also the academic sector is experiencing a significant increase in the demand for identification tools. The option, that data can not only be accessed where they have been created is an essential advantage but this immediately requires a mechanism for identifying those who want to access these data. Wider service, however, results also in a higher risk of being misused. Thus it has to be balanced against increased visibility and vulnerability.

In view of the wide range of the demand for identity services it is not surprising that there is no 'one size fits all' baseball cap solution. Control of access to online banking is on another level than registration for news about upcoming conferences and such differences automatically result in different administrative and operational structures. To these different demands come legal boundary conditions, mostly coupled to national differences. Thus, only a federated approach will allow for the necessary flexibility to balance confidentiality and data privacy against demand for sufficient information and cope with the danger of over-boarding bureaucracy.

On the other hand, the communities requiring these services are highly overlapping. An example is a person, which is interested in online banking is student at a university and part of a team which performs experiments at a large facility. Presently, each service / domain has its own IAA system with own identification standards and users end up with a multitude of different accounts, which is introducing unnecessary complications. That means that bridging between these user federations is very important.

'User friendliness' is one of the key requirements to be observed from the beginning. No solution will be able to define its own operational standard, this standard will be defined by every-day IT tools like Amazon, Facebook, Smart phones Laptop/PC systems. People are no more used to read manuals and will accept only slim, intuitive instantly- and easy-to-use solutions. Everything complicated to use will not survive on the long term.

1.3 Fundamental concepts

1.3.1 Authentication and authorization

An important issue is the definition of authentication and authorization. According to the definition, authentication is the identification of a person by means of specific criteria (e.g. username, password, official document, and picture). Authorization is the way to provide certain rights (e.g. opening a door, accessing a data file) to a person which has been authenticated. Authorization can also be given to groups (e.g. members of a university) but in this document this is understood in the way of a one-to one-relation between person and access right, i.e. that an authenticated person has the role of being part of the university and thus has a certain right. Thus, authentication and authorization are differentiated

but on the other hand, these terms are interrelated, as e.g. a security-critical authorization will in general require a more stringent authentication (e.g. inspection of a passport document instead of a simple Google-type email handshake.

1.3.2 Identity provider and Service provider

Another important functional terms are identity provider (IdP) and service provider (SP). Here an SP is offering a specific service, e.g. an application which requires a license for its use. If a specific user is requesting this service the SP will by means of the IdP determine if he has the right for using this service. That means that the SP sets the rights for its use based on the information from the IdP.

2 Federated Identity Management (FIM) approaches

In the definition and setup phase of a new system definition and development issues naturally play an important role and operational issues are often given secondary priorities. In the case of FIM, however, operation is a key aspect and has to be respected from the beginning. By its nature, the elements of such a service are distributed, in the extreme over the full globe with different national and legal structures and any malfunctions will be very complicated to identify and to cure. FIM is a topic of very general interest for commercial companies as e.g. airline alliances banks, security-sensitive areas over government and public service administrations to practically all academic activities as e.g. universities or research institutions. Clearly, a general treatment would be far beyond the scope of the present project. Therefore, this report will concentrate on FIM related to academic research institutions and her to large research facilities.
A general discussion of federated identity management is the topic of an investigation [1] of the IT situation in a wide range of activities from high-energy physics over climate research, social science and humanities, life science to research at large photon / neutron facilities. Another extended study has been carried through by TERENA [2]. These studies offered an excellent opportunity for positioning the development within WP16 relative to general AAI environment. These studies indicate that the demands from the respective communities are in general very similar. If one, however goes to the details then they are also quite different. The conclusion from that is that a federated solution is the appropriate approach.

2.1 General approach (FIM paper as basis)

A general approach aims at a common vision for FIM across these communities and a common road map towards implementation. By nature of the scientific communities represented, this approach aims for a topology as wide as possible beyond any political or geographical border.
A conceptual example for such an approach is the GRID concept developed in high-energy physics to deal with the vast amount of experimental data from the LHC detectors and compute resources to be distributed to the international

collaboration partners all over the world. Bottlenecks of the GRID approach are complications related to the management of the X.509 certificates and the complicated middleware necessary to manage to access to the data. The goal of the FIM activity [1] is, therefore to arrive at a federated system which is better suited to cope with the present-day demands.
Such a global FIM system could be based on National Research and Education Network (NREN), where the respective members again delegate the further administration to local national institutions. By its concept, the system is as general as possible, there is no relational connection between IdP and SP, any constraint has to be explicitly introduced by e.g. the SP. This system requires trust relations between the various partners and also national legal constraints have to be taken into account. Top goal is to develop an optimal IT solution and science-political aspects are of minor importance. As most of the services envisaged are just starting and frequency of usage is hard to estimate, scalability of any solution is important.

2.2 Dedicated approach

A different approach is taken by the concept developed at the European photon / neutron large facilities. For handling the large number of visiting scientists (more than 30'000 per year), these facilities are running web-based user-office (WUO) systems including local user identification. A specificity of this community is that these facilities are in a mixed cooperation / competition state, comparable e.g. to airline alliances. Here a FIM approach goes for a slim system on top of the existing WUOs, which, by adding FIM functionality, will allow a multitude of novel services (e.g. remote data access, remote experiment access) and at the same time adheres to the confidentiality requirements from the individual facilities.
Characteristic for this approach is that it is not the goal to accept anybody but only users and the science-political environment.
User friendliness is an important issue right from the beginning, as the services envisaged will be used by some of the users only a few times per year. Thus, in addition to controlled access also easy access is a top requirement.
Furthermore, science-political aspects are playing a strong role, as these facility want to stay autonomous and delegate only responsibilities where necessary.
As the local services are in operation since many years and also in view of the new facilities under construction the number of users will not increase strongly (maximum factor 2), scalability is here not so important. In conclusion, the differences to the 'general approach' (2.2) are less computational aspects but management and operation issues.
In order to find within all these boundary conditions the optimal solution with maximum synergy, the photon / neutron community has carried out recently three bi-annual harmonization workshops. The goal was to study the technical possibilities for a community-wide identification system and to determine a solution which at the same time fits into the science-political environment and is in agreement with the operational demands at these facilities. In addition, the solution should be in phase with foreseeable developments of the AAI field.

3 Umbrella as dedicated prototype solution for P / N large facilities

3.1 Requirements

There are several requirements, which have to be met in order for the proposed system to be accepted and put into operation.

3.1.1 Unique user identification

The main goal for a new FIM system is a unique identification of a user on the European scale and beyond. That means that the system must be able to distinguish a 'Peter Fisher' from London University to a person with the same name at Zurich University. This distinction is necessary to be able to uniquely determine if a specific person has access to a certain dataset on a remote server. Present IdPs take care of unique identification within their own set, but uniqueness over several IdPs is not available.

3.1.2 Persistent person-related identity

Present FIM systems often define persons in the form of person@affiliation (Email-type definition). This means, that either a person moving from one affiliation to another (e.g. postdocs) has to change identity or part of the identity is outdated. This is OK for communities which are predominantly static or where there is no need for identity persistence. For users of research facilities, this is not acceptable. The requirement, therefore, that identity has to be related to the person only. . There are approaches [2], which assess to persons unique user IDs. They are, however, far from 100% coverage, which would be necessary for being included in a general solution discussed in the present context.

3.1.3 Low admin level and ease of operation

At present the vast majority of research facilities considered in this context is running since a long time. There are local AAI systems in operation and any new development, in order to be accepted, must be a clear improvement. Administrative resources both, on the user and on the facility side are limited and any increase of overhead is not acceptable. On the facility side, that means, that the load of the new system must be equal or less than that for the existing systems.

On the user side, the load has to be kept as low as possible, e.g. by applying self-service concepts. The strength of use for typical experiments will also vary with time. In the first step, only a proposal is submitted to a facility, which has only a limited chance to be accepted (overbooking). If a proposal is accepted, further administrative steps have to be taken, which may require a higher degree of authentication. Instead of raising the authentication level to a high level from the beginning, a flexible multi-level system may be envisaged requiring only the necessary trust level.

. There is no means of centrally controlling a FIM system and acceptance by the partners involved is to a predominant degree taking place on a voluntary basis. Thus, facility and user friendliness of the system will be decisive for its acceptance.

3.1.4 Confidentiality, data privacy

A good user database is one of the key assets of a user facility and for facility managers it is important that information about 'their' users is not leaking to other facilities. Therefore all aspects of a central database are highly critical.
Concerning forwarding of any user information has to respect the national data protection laws in the various countries. On the other hand, potentially delicate user parameters like gender or age are in part required as part of documents reported to the EU.

3.1.5 Parallel implementation / operation

Implementation of a system for 30'000+ users at two dozen independently running facilities can be performed only in a flexible way. Any time-zero approach would result in a huge administrative load and imply uncontrollable risks. In addition, experiments take years from proposal to publication with the need for accessing all relevant information over the full duration. That means that the system has to be designed such that a parallel use of the old system and implementation of the new one must be possible over years.

3.2 Solution

Based on these requirements the Umbrella system has been developed [3]. First steps up to the production of a prototype have been performed within work package 2 of the EuroFEL ESFRI project [4].

Confidentiality, both in respect to the users of the facilities as also the facilities themselves is a basic requirement for any system dealing with user-related services at the photon / neutron large facilities. That means that authentication and unique user identification is the basic layer of the Umbrella system. Unique user identification needs a certain central component. In order again to comply with the confidentiality requirement, authentication and other elements of the Umbrella are designed such that the central part is kept minimal, just to provide the necessary functionality. Further information needed for the general operation is kept at the local WUOs.

On top of the authentication layer, further tools are implemented, enabling users to easier access information and services at the facilities. They contain basic services such as Account Creator, Attribute Updater, Facility Manager, Module Manager.
The existing Web-Based User Office (WUO) systems at the local facilities contain the whole cycle of handling an experiment, from proposal submission over experiment handling up to the registration of the final publication. In addition, most WUOs include important off-and on-site user services like facility access control or guest-house registration. All these services should be kept and only

missing central services like pan-European user identification should be added (umbrella concept).This requires corresponding modifications for these WUOs to be able to collaborate with and integrate into the umbrella system. Sections 3.3 – 5 will now describe the projected solution in more detail.

3.3 Authentication

The Umbrella authentication system is built upon a user directory where the accounts are stored, and a SAML2 ecosystem, which enables Single Sign-On.

3.3.1 Design

In recent time, there is a vivid development concerning authentication issues both in the academic and commercial sector. An early decision was not to build the Umbrella system completely from scratch but rather build it on top of an existing authentication system. In this way, the project profits from the existing know-how as well as from novel developments.

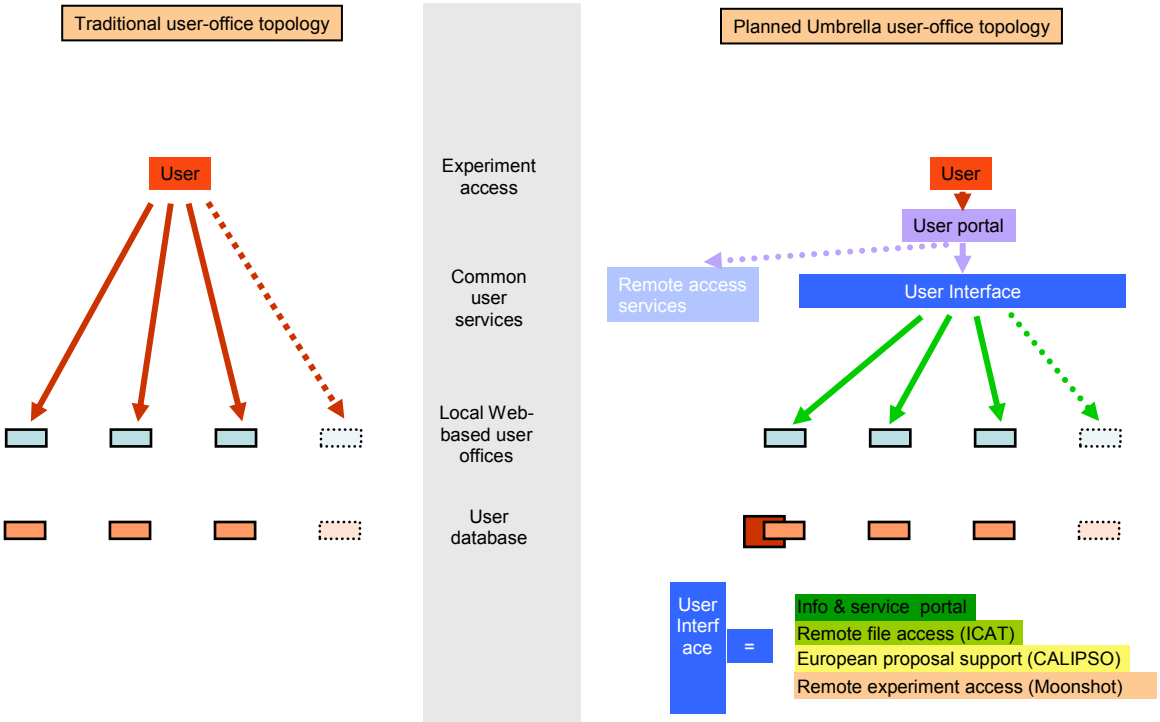

Figure 1, Umbrella topology

3.3.1.1 SAML2 topology

After evaluating different authentication systems, e.g. OpenID, Shibboleth, it was concluded that a SAML2 based system fits the needs best regarding user privacy and data security (see also [5]). SAML2 is an OASIS standard [6] and has wide

industry support; members are: AOL, EMC, Hewlett-Packard, IBM, Microsoft, Nokia, Oracle, Red Hat, SAP, Boeing and others. It has been widely adapted in Switzerland (SWITCHaai), Germany (DFN-AAI), Denmark (DK-AAI) and other countries with hundreds of thousands of users who work with it on a daily base. Finally, it has been decided to use Shibboleth2 [7] by the Internet2 Middleware Initiative as the implementation of the SAML2 specification.

3.3.1.2 Unique user identification

A basic request is *unique user identification*. This can be accomplished only by a central identification element. Thus, Umbrella has only one identity provider, in contrast to other identification systems.

3.3.1.3 Hybrid data structures

Another key requirement is to offer *maximum confidentially concerning users and facilities*. The solution for that is the introduction of hybrid data structures by dividing information in a central part and facility-local parts. For *user information* there is a minimal central part, which contains just as much parameters to necessary to identify a user in a unique way. The rest of the identity information and all authorization information is stored – as it is already now –in the local WUO database. For convenience, e.g. in case of update or migration, the user is able to download, modify and upload his/her local information. *Affiliation information* is also stored in a common database, which is the only possibility, that these data can be used for statistical analyses. Here the relation is inverted, most of the information is stored centrally with the option of storing selected additional items locally.

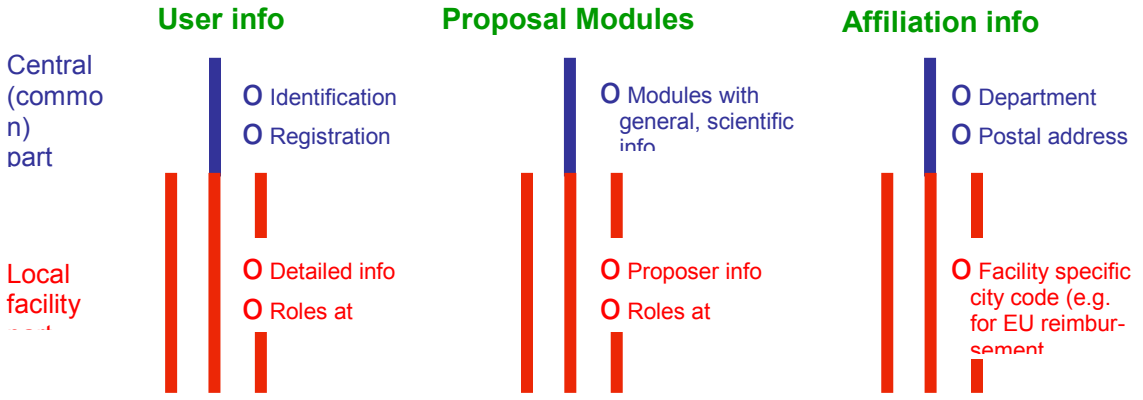

Figure 2, Hybrid data structure

A third dataset is proposals. Proposals play a basic role in the management of experiments at photon / neutron large facilities. Presently, there is no common handling of proposals, which is to a large extent due to the highly confidential

character of these documents. For further extensions the Umbrella concept contains the structure to deal with proposals in a federated way.

3.3.2 Self service

A third key element is *self service*. In order to make the system as slim and efficient as possible, the user enters his/her information by him/herself. As the user is interested in this information being correct, this concept helps to keep the database as up to date as possible. Even more, in this way the legal restrictions can be avoided. In case, user input needs to be validated, this can be accomplished by simple handshake procedures e.g. by involving legal affiliation representatives.
User friendliness is another key element and *multi-level authentication* is an example. Beamlines and experimental stations at large facilities are typically highly overbooked. Thus, the chance of a proposal to be accepted is limited and, in order not to discourage novice users, the threshold for submitting proposals should be low. Similarly, registration for news services is not security-critical. In this case a simple Google-type handshake is a sufficient security concept. This is different from the case, where a user wants to enter restricted locations within a building. Here, depending on specific local security regulations checking of an official personal ID document by a security officer may be required. Such a concept is easily handled by a multi-level authentication system, where the security officers use specific tools to mark the appropriate security levels in the user database entry.

3.3.3 Operation

3.3.3.1 Central part

By its design, there is only a slim central part without need for personal operation. Obviously, a hotline service must be foreseen for covering problems. In addition, it has to be decided, if there will be one central system with practically 100% uptime or if there will be mirrored system with several physical sites.

3.3.3.2 Local WUOs

For the local WUOs operation with Umbrella will not be much different from the present situation. Curation of a local affiliation data base will be no more necessary, as this will be managed centrally. Clearly, there will always be the need for the central database, this will be one of the topics for an operation agreement between then participating facilities.

3.4 Authorization

In principle, authorization is not part of the Umbrella system, as by construction extended user information – i.e. all user parameters beyond the minimal central identification items - and all authorization is stored at the local WUO databases. Nevertheless, minimal central, extended authentication and authorization information belong together – they describe the same person – and, therefore, the format should be compatible between the facilities.

3.4.1 Local parameters

There are numerous parameters, which authorize a user for certain rights by means of ID cards (e.g. badges for beamline access). These parameters are completely of local character and there is no need and also advantage for them to be available to other facilities.

3.4.2 Federated parameters

Different from that are user parameters which control remote access. In principle, the rule should always be that a service provider (SP) decides by itself, if a specific user has the right to access a file at its site and thus this is a SP-local decision. Development and operation, however, are very much simplified, if the corresponding IT structure is standardized.

4 Towards implementation

4.1 Friendly user phase

In the beginning of 2012, the coding of the basic Umbrella components had been completed and it has been decided to ask for the feedback of the final users already at this early stage. For that, clones of operational local WUOs have been linked to the Umbrella and about two dozen 'typical' users from all over Europe and about a dozen facility-IT persons have been asked to play for two months with the system. The overall response was very positive and the concept has been favorable accepted. One of the results was, that the time investment needed for upgrading a typical local WUO to a Umbrella-ready state was of the order of days. This was important information in view of the limited local resources.

4.2 Further software steps

The basic Umbrella system is ready. Nevertheless, several upgrades were necessary before transferring the system from the prototype of the operational status.

4.2.1 Web site

As a web-based tool, Umbrella needs clearly also an own web site. The issue is not so much functional requirements but how this site is related to the web sites of the individual facilities. This topic has, therefore, to be solved in close cooperation of the facilities involved.

4.2.2 Common affiliation database

A common affiliation database for the participating facilities offers several advantages. In the existing local user office systems it is foreseen that not users enter affiliation details by themselves but that this information is entered by the user office staff, as this is the only way the information is stored in such a way that meaningful statistical analyses can be made. This aspect is becoming especially important in view of the increasing number of reports to deliver. The price to pay is that some resource load has to be set aside in order to keep these databases up to date. In case of a common database this load can be shared between the participants. The advantage for the user is, that e.g. in case of a mutation a user registered at several facilities needs to update her/his coordinates only once and the new information is forwarded to all databases.

4.2.3 Logout handling

A typical experience from the 'friendly user' test was the feedback from users that there should be a better logout handling. This is a symptom known from online banking and related to a weakness of the basic internet service. In the meantime, this deficiency could be improved.

4.2.4 Adaptations of Umbrella for remote file access (iCAT)

One of the central novel services to be available with a European identity service is remote file access to experimental user data. During the first years (details to be fixed by the facility data policy) these data are highly confidential, as they are the basis of future publications, the basis of the academic career of the participating researchers. Research teams are only temporarily defined and scientists move between experimental teams. It is, therefore, not sufficient to only assign access rights to persons. The only stable definition is a proposal and which persons have participated in a project defined by a proposal. This information is available in the databases of the local WUOs. That means that the solution is to combine the identity information from Umbrella with the proposal information available at the local WUOs.
A good candidate for remote file access is iCAT, developed at STFC and the next step is to marry the file access concept described above with the iCAT tool.

4.2.5 Adaptations of Umbrella to remote experiment access (Moonshot)

Another attractive application of Umbrella is remote experiment access. Some experiments, e.g. in structural biology, take only hours and long travels for such an experiment do not always make sense. Another case is that senior researchers are often loaded by university teaching obligations and thus not able to participate in an experiment. In such cases it will be very attractive, if these persons can remotely participate in an experiment and have online access to data as they are taken (passive access) or even modify experimental parameters (active access). Access control considerations are very similar to the remote-file-access case mentioned above. An interesting tool candidate is here Moonshot developed at STFC.

4.3 Legal & organizational issues

As part of the transition from the prototype to the implementation stage several minimal administration and legal steps have to be taken. There has to be an agreement of the physical site and topology (single or mirror) of the central server and how – including uptime guarantees - the site is operated and from which resources. The partners have to agree if a formal memorandum of understanding (MoU) is required or if informal mutual agreements are sufficient.
Another issue is managing bodies. Therefore, two teams will be installed,

- (a) a management team including representatives from all partners and dealing with all issues involving resources and authorities of the facilities,
- (b) an technical team dealing with system development issues.

The partners have to agree on security issues, e.g. password standard, setting up a procedure to deal with system break ins.
As user friendliness is of very high importance, user support has to be set up, e.g. hotlining, use of social media.

4.4 Road map

The success of the Umbrella deployment is linked to a realistic road map. The partners have to agree on the priorities of the actions necessary and within which timeframe these should be realized.

Presently, there are several FP7 projects (e.g. CRISP, PaNdata, CALIPSO, NMI3, and Biostruct X) which all deal with user IT-support topics at large European facilities and are in part highly overlapping. In order to avoid any double work or negative competition, all these projects have to cooperate and to adjust their activities. In this respect, the Harmonization meetings (Zürich airport and DESY 2011 and Zürich airport 2012, DESY and HZB 2013) are playing an important role.

Concerning Umbrella, the software developments are described in section 4.2. Sections 4.2.1 and 4.2.3 have been dealt with in spring 2013. Concerning 4.2.2 work is (summer 2013) still ongoing. The prototype is practically ready and work is concentrating on implementation and operational issues, as the work flows of the individual user offices are involved. The other important issue is how to deal with the legal and organizational aspects of Umbrella, especially after the end of PaNdata ODI. Work on a MoU between the partners is ongoing. Significant progress is expected in several meetings in fall 2013.

5 Further developments beyond Umbrella

Work package CRISP/WP16 is concentrating on this aspect. One main issue is non-web-based services (e.g. Moonshot). Another one is bridging, i.e. to connect the photon-neutron federation as covered by Umbrella to other federations like eduGAIN or the HEP federation. Here a first prototype solution has been prepared within WP16. More detailed descriptions are available from the related CRISP documents and deliverables [7].

6 Concluding remarks

The goal of CRISP / WP16 [present work] and PaNdata / WP3 [8] is to provide the users of the European large research facilities with special emphasis on photon / neutron facilities with a federated authentication and authorization system. As solution the Photon / Neutron facilities have selected the Umbrella system as a slim layer on top of the local user office systems (DUO, SMIS, and VUO) installed at the local facilities. For the photon / neutron facilities the important functionalities are available and, therefore, PaNdata / WP3 concentrates on the implementation of Umbrella at these facilities. See in a wider context, there are other academic and commercial federated-identity activities and it is important to study the respective bridging aspects and to explore position and evolve the Umbrella solution in this context. This will be the central contribution of CRISP / WP16.

7 References

[1] *Federated Identity Management for Research Collaborations*, B. Jones et al. 2012
[2] Advancing technologies and Federating communities, Study on Authentication, Authorization and Accounting (AAA) Platforms For Scientific data / information Resources in Europe, TERENA 2012 <https://confluence.terena.org/display/aaastudy/AAA+Study+Home+Page
[3] *Functional Description of the EUU / UAA Tools,* R. Dimper, D. Porte, O. Schwarzkopf, D. Feichtinger, D. Lauk, H.J. Weyer, 2010
[4] The Umbrella System, Bjoern Abt, Heinz j Weyer, deliverable 2.5, EuroFEL, 2011
[5] <http://identitymeme.org/doc/draft-hodges-saml-openid-compare.html
[6] <http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security>
[7] <http://www.switch.aai/index.html>
[8] PaNdata deliverable D3.1 Specification of AAA infrastructure
[9] CRISP deliverable D16.1 Specification of AAI infrastructure
[10] <http://about.orcid.org/>